NUI Galway
OÉ Gaillimh

Autumn Examinations 2014

Exam Code(s)	4BCT
Exam(s)	4th B.Sc. (Information Technology and Computer Science)
Module Code(s)	CT422
Module(s)	Modern Information Management
Paper No.	
External Examiner(s)	Dr. John Power
Internal Examiner(s)	Prof. G. Lyons Dr. M. Madden Dr. C. O'Riordan
Instructions:	Answer any 3 questions
Duration	2 hours
No. of Pages	3 (including cover page)
Discipline(s)	Information Technology
Requirements	None

1. (a) Describe the vector space model approach to Information Retrieval. Your answer should include a description of the query and document representations and the comparison approach used. (12)

   (b) In the vector space model, terms are considered to be independent of each other. Outline how you might calculate or estimate the degree of relatedness between terms occurring in a collection. (12)

   (c) Prior to calculating the similarity between a query and a document, documents and queries are often pre-processed. Common techniques include *stemming* and *stop word* removal. Explain these processes and the possible effects they may have on the retrieval process. (9)

2. (a) Many modern web-based search engines attempt to take into account the web link structure in addition to the content of the pages. Describe briefly the Page Rank algorithm that uses information embedded in the web link structure to return relevant documents to a user. Include an example to illustrate your answer. (14)

   (b) What is meant by *collaborative filtering*. Illustrate how a collaborative filtering approach could be used to make a prediction regarding *The Incredibles* for user *Ciara* given the following data set: (10)

	Up	*The Incredibles*	*Peter Pan*	*Toy Story*
Ciara	4		3	2
Muireann	1	4	3	
Daria	3	3		
Mia	4	5	4	3

   (c) Discuss what you consider to be the main limitations of a collaborative filtering approach and suggest approaches to overcome these limitations. (9)

3. (a) Empirical evaluation of information retrieval systems plays an important role in information retrieval research. Define and discuss the following metrics that can be used to measure the performance of an Information Retrieval system: *precision, recall, novelty* and *coverage.* (11)

   (b) Suggest a suitable compression algorithm to deal with large document collections suitable in the domain of Information retrieval. Your answer should include a discussion of approaches to compress indexes. (11)

   (c) Describe and discuss, with the aid of examples, suitable indexing strategies and algorithms to deal with single term queries and prefix queries. (11)

4. (a) Supervised learning approaches have been adopted in information retrieval systems to either adapt to changes in user behaviours or to learn an optimal manner in which to combine information or process information to give good performance. Discuss any learning approach in relation to a problem of your choice in information retrieval. Your answer should also identify the strengths and weaknesses of this approach. (11)

   (b) Query modification is often used by systems to attempt to improve precision and recall for a given information need. Discuss an approach that uses user feedback on the returned answer to modify the query. Illustrate the approach with a suitable example. (11)

   (c) Traditionally, information retrieval systems and web search engines have presented results in a list ordered by estimated relevance. Identify any limitations with this approach and suggest an alternative presentation of results. (11)